OPEN DATA AND CHALLENGES FACED BY NATIONAL STATISTICS OFFICE

Siti Haslinda Mohd Din, Nur Aziha Mansor, Faiza Rusrianti Tajul Arus
Department of Statistics, Malaysia

An amazing growth of interest in open data worldwide over the past half-dozen years has derived tremendous open data initiatives (ODI). Open data of which defined as data that can be freely used, shared and built-on by anyone, anywhere, for any purpose. The benefits of the availability of open data can be served many groups including government. Open data creates value for the government. It is able to improve measurement of policies, better government efficiency, deeper analytical insights, greater citizen participation, and a boost to local companies by way of products and services that use government data. However, in order to gain these huge potential benefits, the question of what is hindering the success of open data and what are the real challenges need to be worked on. The challenges that need to be considered include the standard guides and definitions of open data; the suitable technologies for opening up data; technological barriers for governments; no structured approach to openness; too little research on the impact of openness; the privacy issues; and the mosaic effect.

Keyword: Open data, National Statistics Office, Open data initiatives

1. INTRODUCTION

Information is a tool for supporting development, knowledge sharing and social initiatives. With the vast quantities of data and this data can be transform virtually to enhance the quality of living. It makes precise the meaning of “open” in the terms “open data” and thereby ensures quality and encourages compatibility between different pools of open material. Open data are can be freely used, modified, and shared by anyone for any purpose.

Open data has potential to create tremendous value and has started to be used on a wider scale. New products and business models are emerging off the back of the Open Data movement. Hence, by using App developers the weather reports to warn people of pollution in specific areas can be made. Traffic data is being used for real-time traffic reporting to ease congestion in urban areas. Government data is being utilised to track how tax income is being spent. Open data is helping people improve their household energy efficiency and linking property owners with construction companies that can make it happen. There are many great examples of how open data is already saving lives and changing the way we live and work.

“Open Data” is generally understood to mean data that are made available to the public free of charge, without registration or restrictive licenses, for any purpose whatsoever (including commercial purposes), in electronic, machine-readable formats that are easy to find, download and use. As applied to public institutions such as governments and intergovernmental organizations, Open Data is grounded in the recognition that government data is produced with public funds so, with few exceptions, should be treated as public goods.

Data that are publicly available do not necessarily meet the definition of “Open Data.” For example, data from NSOs may be publicly available, but only to certain qualified or registered users, or with narrow restrictions on how the data can be used. Data may also be publicly available but only in proprietary formats that are difficult to access or manipulate (such as PDF), or even non-electronic formats. Nonetheless, these data are often characterized as “public data.” It is important for NSOs to clearly differentiate their data products and corresponding policies.

2. CHARACTERISTICS OF OPEN DATA

Key characteristic of Open Data is the potential for reusability, both by data experts and the public at large. Reusability is the key to creating new opportunities and benefits from government data as detailed later in this working paper. For Open Data to be reusable it must generally meet two basic criteria. First, the data must be *legally* open, meaning that it is placed in the public domain or under liberal terms of use with minimal restrictions. This ensures that government policies do not create barriers or ambiguities concerning how the data may be used. Second, the data must be *technically* open, meaning that it is published in electronic formats that are machine-readable and preferably non-proprietary. This ensures that ordinary citizens can access and utilize the data with little or no cost using common software tools.

Open data definition gives precise detail of summaries the characteristics of open data are:

- Availability and Access: the data must be available with no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
- Universal Participation: must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.
- Re-use and Redistribution: the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.

Government data shall be considered open if it is made public in a way that complies with the principles below:

1. Complete - All public data is made available. Public data is data that is not subject to valid privacy, security or privilege limitations.
2. Primary - Data is as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.

3. Timely - Data is made available as quickly as necessary to preserve the value of the data.
4. Accessible - Data is available to the widest range of users for the widest range of purposes.
5. Processable Machine - Data is reasonably structured to allow automated processing.
6. Non-discriminatory - Data is available to anyone, with no requirement of registration.
7. Non-proprietary - Data is available in a format over which no entity has exclusive control.
8. License-free - Data is not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed.

3. OPEN DATA INITIATIVES BY NSOs

Open Data initiatives are transforming how governments and other public institutions interact and provide services to their constituents. They increase transparency and value to citizens, reduce inefficiencies and barriers to information, enable data-driven applications that improve public service delivery, and provide public data that can stimulate innovative business opportunities.

National Statistics Offices (NSOs) produce many datasets that could typically comprise the foundation of an Open Data program. With relationships with other data producing agencies and expertise in dealing with technical and data quality issues, NSOs are extremely well placed to make a valuable contribution to Open Data initiatives.

Despite these advantages, NSOs do not always feature conspicuously in government sponsored Open Data programmes and may be missing an important opportunity to expand the use and re-use of the data.

3.1 Open Data initiatives have significant implications for NSO operations.

NSOs manage many surveys and census such as economic survey and household census for gather economic statistics as well as demographic data series, which are considered essential, high-value datasets in Open Data programs; hence, NSO products and expertise will be in high demand. The underlying principles of Open Data—to make government data as accessible and useful as possible—are clearly aligned with the NSOs core mission. Some requirements of Open Data are likely to be easily achievable by NSOs, such as opening data that is already public, and allowing data that is already available in non-open formats to be downloaded in bulk formats. Other desirable features of a good Open Data initiative, such as the transparent publishing of relevant policies, metadata, and training materials, will likely take more effort.

3.2 NSOs have strong roles to play in Open Data initiatives

While NSOs are not usually well positioned to lead a government-wide Open Data initiative, they nonetheless can be a vital component of it. NSOs produce high-demand official statistics, and can be instrumental in ensuring that the Open Data initiative is properly aligned with the wider National Statistical System. NSOs have extensive experience in data selection, the application of standards, and provision of metadata; hence, NSOs have a clear role to play in providing guidance to other agencies in publishing their own data. Finally, as members of the international statistics community, NSOs are in a good position to make sure that the Open Data initiative is well aligned with the efforts of other countries and international organisations, and in line with internationally agreed standards.

3.3 Open Data initiatives can greatly benefit NSOs

Data collection is a costly activity, financed by public spending and other investments. Open Data can also relieve pressure on NSO operating budgets by reducing the demand for custom tabulations and other data requests. Open Data will likely raise the profile of NSOs both within the government, with key agencies and the public. A greater profile will strengthen the NSOs reputation and opportunities.

Data quality is often cited as a concern preventing NSOs from adopting Open Data policies. Statistics are subject to various quality factors which are inherent to standard statistical methods. The important thing for NSOs is to provide adequate context and information about the uncertainties and limitations of any particular dataset. Doing so would lead to increased transparency and greater public trust in both the data and the NSO itself. Furthermore, NSOs can use feedback mechanisms available through Open Data to improve data quality.

3.4 Open Data raises potential challenges for NSOs

NSOs have the responsibility, often established in legislation, to protect the confidentiality and privacy of their data providers, who may be individuals, households or businesses. Confidentiality issues do not apply equally to all types of data, and many types of data, such as aggregated statistics, can typically be published and opened without breaching confidentiality as standard anonymization techniques are applied. In the more sensitive realm of survey microdata, each NSO must ultimately make the determination about whether and how to make these data public and which techniques to be use, but experience to date includes several cases where microdata have been opened without compromising confidentiality.

Open Data best practices require that data producers provide clear terms of use with minimal restrictions on how data can be used. This may prove challenging for NSOs that manage different classifications of products, some of which may be restricted on grounds of confidentiality or even national security. However, NSOs can take advantage of standard international licenses, and there are several case studies for managing data under multiple access policies.

NSOs may also be concerned about the resources and capacity required to implement Open Data. Experience to date suggests that the additional resources are not substantial, and may be at least partly offset by cost savings and greater efficiencies. But it is true that Open Data may represent an opportunity costs to some NSOs, to the extent that they derive some revenue from data sales.

NSOs may also face challenges in engaging a larger user ecosystem. Since the goal of Open Data is to increase the use of data, it is almost inevitable that the NSO will

be interacting with user communities which are new and possibly unanticipated. Again, this can be an opportunity for the NSO as much as a challenge. Many public agencies that have adopted Open Data policies have experimented using workshops, toolkits, competitions, social media and other approaches to successfully engage new data users. Ultimately, what may be required is for NSOs to simply be open to change, in order to reap the benefits that Open Data can provide them.

4. CHALLENGES FACES BY NSOs

As we know, open data represents a new era in data exploration and utilization but it can be fraught with challenges. The hindering factors that affect the success of open data and what are the real challenges need to be worked on. Government as well as NSOs truly believed in the opportunities of open data but the use of open is accompanied by many impediments. This paper highlighted several challenges that need to be identified by NSOs in utilization of open data.

4.1 Standard guides and definitions of open data

Government and NSOs should establish the standard guides and definition of what data should be public and how to make data public. The data openness means the NSOs will make public information available proactively and put the information within the reach of the public (online) without barriers for its reuse and consumption. Strong open data policies should be built upon the principles embodied by existing laws and policies that defend and establish public access, often defining standards for information quality, disclosure and publishing (Sunlight Foundation, 2014).

4.2 Suitable technologies

It can happen where some local governments or NSOs cannot identify the relevant for opening up data and therefore they fail to implement proper technologies. They are unable to ascertain relevant applications for their data. Besides that, the data also should be available in a machine-readable standard format, which means it can be retrieved and meaningfully processed by a computer application. According to Emily Shaw, the national policy manager of Sunlight Foundation, if the intended users are

developers and programmers the data should be presented within an application programming interface (API); if it's intended for researchers in academia, data might be structured in a bulk download; and if it's aimed at the average citizen, data should be available without requiring software purchases (Shuen, 2014).

4.3 Technological barriers

Basic technological components for open data are still not easily accessible to smaller governments or NSOs. NSOs might have limited independence and unstable budgets as well as lack of local technical skills. As discussed in Istanbul by 15 open data and transparency leaders from across Eastern Europe and Central Asia organized by UNDP and the Partnership for Open Data (the World Bank, the Open Data Institute and Open Knowledge) in September 2014, in order to know how much government should expect to spend on an open data initiative, cost estimates are typically required for:

- Allocating budget(s) for the central team and within individual ministries, departments or agencies.
- Budgets need to reflect the additional and specific resources required for a project; when further capacity was required.
- A business case for the open data initiative.

Besides that, communication with developers and technical peoples on open data also should be part of open data strategy in order to reduce technological barriers.

4.4 Structured approach to openness

In the age of open data, all public data should be made available where the data is not subject to valid privacy, security or privilege limitations. The data supposes as it is collected at the source, with the highest possible level of granularity, not in aggregate or modified forms. Therefore, the data should reasonably structure to allow automated processing. The ability for data to be widely used requires that the data be properly encoded. Free-form text is not a substitute for tabular and normalized records. Images of text are not a substitute for the text itself. Sufficient documentation on the data format and meanings of normalized data items must be available to users

of the data. This principle was agreed by 30 open government advocates during the meeting in Sebastopol, California, December, 2007 (Tauberer, 2007).

4.5 Research on the impact of openness

There are few studies done by academics (instead of consultants) on the impact of data openness. The studies often take a macro economic perspective and do not assess micro levels of impacts. In general, public had mixed views as to whether trust could be built by opening up data or whether trust could be undermined if confusion was caused by multiple interpretations of the same data. Thus, more researches in this area should be intensified to prove that making information accessible to the public can improve public service delivery and lowering the cost of getting information.

4.6 Privacy issues

In reality, government agencies collect huge amounts of data about businesses and households. Technological advances have made those data possible to be accessed easily. Therefore, governments as well as NSOs should alert to what extent privacy concerns should be handled at technological level advances. Key concern is the sensitivity of the data in which privacy rights might be damaged. However for transparency and open data supporters, ‘data protection’, or ‘protecting privacy’ might be used as excuses not to release data, or to only release data in aggregated forms that don’t permit detailed analysis of what government is doing (Open Data Research Network, 2013).

4.7 Mosaic effect

“According to Marion Royal, program director of U.S. Data.gov, when very large data sets, even those with completely unclassified information, are combined. People can mix that data, reassembling it in unforeseen ways, like a mosaic puzzle. In a worst case scenario, security can be compromised by those with ill intent” (Breeden, 2014). This is the phenomenon where non-personally identifiable information can be

combined with other available information in such a way as to pose a risk of identifying an individual.

5. CONCLUSION

The use of open data is a recent phenomenon but, as with many technological advances, it is growing in relevance and prevalence to become “normal.” Although clearly a good thing in theory, open data risks and challenges must be given equally concern. Governments need to protect individuals and organizations from the risks of open data and at the same time advancing open data’s potential value. The risk of utilization of open data is there unless approached right. With overloaded information, NSOs need to find better ways to compile, process, store and share the “sea of information” without compromising quality and confidentiality. The statistics can typically be published and opened without breaching confidentiality so long as standard anonymization techniques are applied. The new data policy that opens public datasets to all users free of charge or in other word, increasing access to information can be an opportunity to NSOs to improve the data quality. It aligns with “good government” efforts, in which publicly available information is seen as essential to transparency, public trust, and improving public services. NSOs should transform the challenges faced into opportunities and try to take advantages of open data. The new abilities bring by open data can eliminate the loose procedure in data collection, compilation and processing in order to enhance the quality of data. In the nutshell, open data gives excuses to NSOs make changes.

6. REFERENCES

10 challenges for open data. (2015, August). Retrieved from www.openstate.eu/en/2015/08/english-10-challenges-for-open-data/

Breeden, J. (2014, March). Worried about security? Beware the mosaic effect. *GCN*. Retrieved from https://gcn.com/articles/2014/05/14/fose-mosaic-effect.aspx

Guidelines for open data policies. (2014, March). *Sunlight Foundation*, *3*. Retrieved from http://assets.sunlightfoundation.com/policy/Open%20Data%20Policy%20Guidelines/OpenDataGuidelines_v3.pdf

How to plan and budget an open data initiative. (2014). Retrieved from https://theodi.org/guides/how-to-plan-and-budget-an-open-data-initiative

Open data & privacy. (2013, August). Retrieved from http://www.opendataresearch.org/content/2013/501/open-data-privacy-discussion-notes

Parker, B.J. & Jain, K. (2015, April). The challenges of open data and privacy issues. *Western City.* Retrieved from http://www.westerncity.com/Western-City/April-2015/The-Challenges-of-Open-Data-and-Privacy-Issues/

Shuen, J. (2014, March). Open data: what is it and why should you care? Retrieved from http://www.govtech.com/data/Got-Data-Make-it-Open-Data-with-These-Tips.html

Tauberer, J. (2007) 8 principles of open government data. Retrieved from https://opengovdata.org/

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. Information Systems Management, 29(4), 258-268.